



Original Articles

Fluid intelligence and working memory support dissociable aspects of learning by physical but not observational practice

Dace Apšalka^{a,b}, Emily S. Cross^{a,c,*}, Richard Ramsey^{a,*}^a Social Brain in Action Laboratory, Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Wales, UK^b MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, England, UK^c Institute of Neuroscience and Psychology, School of Psychology, University of Glasgow, Glasgow, Scotland, UK

ARTICLE INFO

Keywords:

Observational learning
Individual differences
Working memory
Fluid intelligence

ABSTRACT

Humans have a remarkable ability to learn by watching others, whether learning to tie an elaborate knot or play the piano. However, the mechanisms that translate visual input into motor skill execution remain unclear. It has been proposed that common cognitive and neural mechanisms underpin learning motor skills by physical and observational practice. Here we provide a novel test of the common mechanism hypothesis by testing the extent to which certain individual differences predict observational as well as physical learning. Participants ($N = 92$ per group) either physically practiced a five-element key-press sequence or watched videos of similar sequences before physically performing trained and untrained sequences in a test phase. We also measured cognitive abilities across participants that have previously been associated with rates of learning, including working memory and fluid intelligence. Our findings show that individual differences in working memory and fluid intelligence predict improvements in dissociable aspects of motor learning following physical practice, but not observational practice. Working memory predicts general learning gains from pre- to post-test that generalise to untrained sequences, whereas fluid intelligence predicts sequence-specific gains that are tied to trained sequences. However, neither working memory nor fluid intelligence predict training gains following observational learning. Therefore, these results suggest limits to the shared mechanism hypothesis of physical and observational learning. Indeed, models of observational learning need updating to reflect the extent to which such learning is based on shared as well as distinct processes compared to physical learning. We suggest that such differences could reflect the more intentional nature of learning during physical compared to observational practice, which relies to a greater extent on higher-order cognitive resources such as working memory and fluid intelligence.

1. Introduction

A remarkable feature of human cognition is the ability to acquire skills through passive observation. Whether learning to tie shoelaces or dance “the robot”, one can acquire complex skills by physically practicing them or by watching others perform them. Recent proposals have suggested that learning by physical and observational practice relies on a common set of cognitive and neural mechanisms. To date, however, research is only beginning to elucidate the ways in which both types of learning draw upon on shared mechanisms. One way to deepen understanding of shared mechanisms between physical and visual learning, which has not been used previously, is through demonstration of common individual differences that influence both kinds of learning. Therefore, the current study tested the extent to which individual differences in cognitive abilities predict learning following physical and

observational practice.

Common cognitive and neural mechanisms have been previously associated with action perception and production (Caspers, Zilles, Laird, & Eickhoff, 2010; Prinz, 1997; Rizzolatti & Sinigaglia, 2010, 2016). Hence, simply watching actions performed by others engages visuomotor representations, which are also engaged during the performance of similar actions (Gentsch, Weber, Synofzik, Vosgerau, & Schütz-Bosbach, 2016). In addition, common cognitive and neural mechanisms have been associated with learning motor skills by physical and observational forms of practice (Hodges, Williams, Hayes, & Breslin, 2007; Vogt & Thomaschke, 2007). For example, neuroimaging studies have shown that similar frontoparietal brain regions are associated with physical and observational practice (Cross, Kraemer, Hamilton, Kelley, & Grafton, 2009; Higuchi, Holle, Roberts, Eickhoff, & Vogt, 2012; Kirsch & Cross, 2015; Sakreida et al., 2018; Vogt &

* Corresponding authors at: Social Brain in Action Laboratory, Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Wales, UK.
E-mail addresses: emily.cross@glasgow.ac.uk (E.S. Cross), r.ramsey@bangor.ac.uk (R. Ramsey).

Thomaschke, 2007). Moreover, occupying the motor system with another task (Mattar & Gribble, 2005) and applying neurostimulation to sensorimotor cortices (Brown, Wilson, & Gribble, 2009; McGregor, Cashaback, & Gribble, 2016), both impair observational learning. Finally, behavioural measures of performance have been demonstrated to be influenced in a similar manner following observational and physical forms of motor learning (Bird, Osman, Saggerson, & Heyes, 2005; Blandin, Lhuisset, & Proteau, 1999; Boutin, Fries, Panzer, Shea, & Blandin, 2010). Together, these studies have provided evidence at both a cognitive and a neural level that motor skill acquisition via physical and observational practice partly rely on common mechanisms.

Although evidence suggests that common mechanisms operate in different forms of motor learning, current understanding of observational learning remains in its infancy and is therefore guided by a relatively impoverished set of neurocognitive models. Such models stipulate that shared processes are likely to be implemented in frontoparietal brain circuits (Cross et al., 2009; Gardner, Aglinskas, & Cross, 2017; Higuchi et al., 2012; Lago-Rodríguez & Cheeran, 2014; Mattar & Gribble, 2005). However, beyond identifying the neural circuits that are involved, there remains a relatively coarse understanding of the type of cognitive mechanisms that are shared, due to a lack of research that characterises the structure of such cognitive systems. As such, a novel way to study the extent to which motor learning through physical and observational practice relies on shared cognitive mechanisms is to probe learning rates as a function of common individual differences. In the context of motor learning, studies have investigated individual differences following physical practice (Ackerman & Cianciolo, 2000; Ackerman, 1988), but to our knowledge, no studies have investigated individual differences in motor learning through observational practice. Therefore, the extent to which common cognitive abilities predict performance during both types of learning is currently unclear. By focussing on individual differences, we are able to develop a deeper understanding of the structure of the cognitive systems that are involved in observational learning, as well as the extent to which these systems operate in the same manner when acquiring motor skills through physical learning.

Individual difference research in the context of learning has a long history due to its potential to inform instruction in educational settings (Gagné, 1967; Jonassen & Grabowski, 2012). General cognitive abilities, such as working memory and fluid intelligence, have long been studied in relation to learning. Working memory has been characterised as temporary storage and manipulation of information relevant for performance in cognitive tasks (Baddeley, 1992). In contrast, fluid intelligence refers to an ability to reason and is typically measured using tasks that require the solving of novel problems (Cattell, 1971; Horn, 1976). Although correlated, it has been demonstrated that working memory and fluid intelligence are partly dissociable constructs and may be associated with different components of learning (Ackerman, Beier, & Boyle, 2005; Kane, Hambrick, & Conway, 2005; Shipstead, Harrison, & Engle, 2016). For example, using a rule-learning task, fluid intelligence has been shown to predict learning and retrieval processes above and beyond working memory (Wang, Ren, & Schweizer, 2017).

In terms of motor learning through physical practice, evidence to date suggests that working memory predicts future gains in skill acquisition (Bo & Seidler, 2009; Unsworth & Engle, 2005). For example, using a motor sequence learning task, Bo and Seidler (2009) demonstrated that greater working memory capacity was associated with the ability to “chunk” sequences into longer components and faster rates of learning. However, it is unclear if working memory and fluid intelligence predict dissociable components of motor learning in a similar manner to other forms of rule-based learning (Wang et al., 2017). Indeed, if motor learning follows findings from prior rule-based learning studies of individual differences (Wang et al., 2017), working memory and fluid intelligence should have dissociable influences on performance. Working memory should be associated with more general dimensions of learning, such as task preparation and attention, whereas

fluid intelligence should be associated with cognitive operations that are tied to action sequences more specifically, such as memory retrieval and specification of sequence-specific information. The current study design enables the distinction between general learning and sequence-specific learning by employing a pre- and post-test, as well as focusing on performance of trained and untrained sequences. That is, sequence-specific learning is indexed by differences between trained and untrained sequences at post-test, whereas general learning is indexed by improvements from pre- to post-test that generalise to untrained sequences.

In contrast to physical practice, no research to date has investigated how cognitive abilities such as working memory and fluid intelligence relate to skill development through observational practice. Although the characteristics of the observer have been argued to influence the extent to which observational learning occurs (Bandura & Walters, 1963), very little is known about individual differences in learning through observational practice. Indeed, there is some suggestive evidence that individual differences in observational learning operate in educational settings (Koran, Snow, & McDonald, 1971), but no research in the domain of motor learning through observation has examined these questions directly.

The aim of the current study is to bridge this gap by investigating motor learning through physical and observational practice by studying individual differences. Support for a common process account of motor learning would be provided if working memory and fluid intelligence similarly predict learning gains following physical and observational practice. In contrast, limits to the common process account would be revealed if distinctly different patterns emerge in terms of the relationship between cognitive abilities and learning following physical and observational practice. Support for either model of underlying processes, however, would enrich current understanding of the cognitive basis of observational learning and the extent to which shared mechanisms operate in distinct learning contexts.

In addition to testing these primary research questions, which had clearly developed hypotheses, we also designed the study to enable a set of exploratory analyses to be performed. The exploratory part of the study focussed on understanding the relationships between components of personality and learning through physical and observational learning. That is, we were interested in exploring the degree to which personality measures such as empathy, interdependence, narcissism and Big-Five personality dimensions would predict learning rates, as it is conceivable that the degree to which individuals are broadly self or other focussed may impact skill acquisition in the learning contexts under investigation. For example, one may expect that personality profiles with a greater focus on others, such as higher empathy, higher interdependence, and lower narcissism, may learn more from others during observational learning. In contrast, individuals with a personality profile with a greater self-focus may benefit learning through physical more than observational practice.

2. Method

Consistent with recent proposals (Simmons, Nelson, & Simonsohn, 2011, 2012), we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. In addition, following open science initiatives (Munafò et al., 2017), the data, stimuli and analysis code associated with this study are freely available online (osf.io/n8sqv/). By making the data available, we enable others to pursue tests of alternative hypotheses, as well as more exploratory analyses.

2.1. Participants

We determined our sample size by terminating data collection once we had collected over at least 100 participant datasets per group as this permitted sensitivity to detect at least small-to-medium effects (for

Table 1
Participant characteristics.

	Physical practice (N = 92)	Observational practice (N = 92)	Group difference (p-value)
Demographics			
Gender (male:female)	30:62	30:62	1
Age (years; M ± SD)	19.68 ± 2.32	19.70 ± 2.62	0.976
English 1st language (yes:no)	76:16	80:12	0.538
Baseline performance (M ± SD)			
Pre-Execution time (s)	2.16 ± 0.78	2.00 ± 0.57	0.113
Personality measures (M ± SD)			
Working memory	196.13 ± 15.75	195.18 ± 15.17	0.679
Fluid intelligence	31.03 ± 8.01	31.78 ± 7.22	0.506
Extraversion	3.32 ± 0.78	3.2 ± 0.73	0.227
Agreeableness	3.76 ± 0.67	3.77 ± 0.6	0.915
Conscientiousness	3.23 ± 0.55	3.29 ± 0.65	0.668
Neuroticism	3.14 ± 0.8	3.24 ± 0.84	0.487
Openness	3.4 ± 0.6	3.46 ± 0.59	0.359
Perspective taking	18.88 ± 4.47	17.96 ± 5.35	0.371
Fantasy	18.67 ± 6.01	18.74 ± 6.08	0.992
Emotional concern	19.03 ± 4.56	19.35 ± 5.34	0.426
Personal distress	12.36 ± 4.8	12.7 ± 5.23	0.616
Narcissism	12.41 ± 6.27	11.6 ± 6.88	0.261
Interdependence	43.3 ± 5.83	43.03 ± 6.9	0.917

details on our power analyses, see the *Hypothesis testing* section below). Two hundred and twenty-three Bangor University student volunteers took part in the study: 69 males and 154 females, 18–37 years old ($M = 19.96$ years, $SD = 3.09$). All but one participant were right-handed (based on self-report). The left-handed participant was excluded from the sample. Data of additional 38 participants were also excluded for reasons provided below (see *Data processing*). The final sample comprised 184 participants. Participants were randomly assigned to physical ($N = 92$) or observational ($N = 92$) practice groups. There were no significant differences between the two groups in terms of demographics and baseline performance (Table 1). Participants provided their written informed consent prior to beginning all experimental procedures. Participation was rewarded with course credits or £10. The study was conducted in accordance with the Declaration of Helsinki and all procedures were approved by the Ethics Committee of the School of Psychology at Bangor University (approval number: 2014-11824) and the UK Ministry of Defence Research Ethics Committee (approval number: 524/MODREC/14).

2.2. Measures of individual differences

2.2.1. Fluid intelligence

Fluid intelligence was assessed by a total score of the Analogies, the Number series and the Matrices subtests of The Intelligenz-Struktur-Test 2000R (Amthauer, Broeke, Liepmann, & Beauducel, 2001), as applied before by Beauducel, Broeke, and Liepmann (2001). A computerised version of the subtests was created in MATLAB 8.3.0 (The MathWorks, MA, USA), closely mimicking the paper version of the tests.

The Analogies subtest measures the ability to reason and see relationships between words. For example, presented with question ‘dark : light = wet : ?’ and a list of five possible answers ‘rain’, ‘day’, ‘damp’, ‘wind’, ‘dry’, participants would select one that completes the given analogy. In this example, the correct choice would be ‘dry’, because ‘dry’ is the opposite of ‘wet’, as ‘light’ is the opposite of ‘dark’.

The Number series subtest assesses numeric reasoning ability. Participants are asked to find the number that is next in the line of a given numerical sequence that is built up according to a specific rule. For example, given a sequence ‘2 4 6 8 10 12 14 ?’, participants would

need to infer that in this sequence every number that follows is by two greater than the one before, as such, the next number that follows would be 16.

The Matrices subtest assesses abstract, figural reasoning ability. Participants are presented with three matrices with different shapes or content. The task is to identify a specific pattern which links the three matrices and from a set of five alternatives to choose the fourth matrix which follows the same common pattern. Each of the three subtests contains 20 items and participants have to complete as many items as possible in the given time limit (7 min for Analogies and 10 min for the Number series and for the Matrices). Participants were awarded 1 point for each correct answer and the results of the three subtests were summed together to obtain the fluid intelligence score.

2.2.2. Working memory

Working memory was assessed by a computerised version of the spatial short-term memory test, implemented and validated by Lewandowsky, Oberauer, Yang, and Ecker (2010). Although there are many different types of working memory task, we chose one that was primarily visuospatial to be consistent with the main task, which has primarily visuospatial features. In brief, participants had to remember spatial relations between dots in a 10×10 grid. Two to six dots were presented, one by one, for 900 ms each, with an intertrial interval of 100 ms. After all of the dots were shown, participants were asked to remember the presented pattern of dots by clicking the cells in an empty grid. The order and the absolute position of the dots were irrelevant, only the overall pattern had to be recalled. There were 30 trials in total, 6 at each set size. The order of trials and dot sequences was the same for all participants.

The working memory score was calculated based on the similarity between the presented and recalled patterns (Lewandowsky et al., 2010). For each dot, two points were awarded if participants clicked on the exact location, one point was awarded if they clicked within one grid place from the exact location, and zero points were awarded if they clicked more than one grid place from the exact location. The total score was the sum of all scores on all trials with the maximum possible score being 240.

2.2.3. Personality questionnaires

To support exploratory analyses, we used multifaceted empathy, interdependence, narcissism and Big-Five personality measures to assess individuals’ self-other relations and broad personality characteristics. Empathy scores were acquired using the interpersonal reactivity index questionnaire (IRI; Davis, 1980, 1983). The IRI is a 28-item measure of four empathy dimensions: perspective taking (adopting other’s point of view), fantasy (self-identification with fictional characters), empathic concern (compassion and concern for others), and personal distress (distress when seeing another’s negative experience). Interdependence was assessed by a 24-item Self-Construal scale (Singelis, 1994). The scale measures both interdependence and independence, but in the analysis, we focused only on the interdependence measure. Trait narcissism was measured by a 40-item Narcissistic personality inventory (NPI; Raskin & Terry, 1988). Broad personality characteristics were assessed by a 44-item Big-Five inventory (John, Donahue, & Kentle, 1991; John, Naumann, & Soto, 2008) measuring five domains of personality: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. All four questionnaires were created in MATLAB and required forced-choice responses.

2.3. Stimuli

A keypress sequence-learning paradigm was used, which the authors have used previously with neurostimulation and neuroimaging methods (Apšvalka, Cross, & Ramsey, 2018; Apšvalka, Ramsey, & Cross, 2018). The paradigm was based on the task used by Wiestler and

Diedrichsen (2013). A standard QWERTY black computer keyboard had the Q, 3, 4, 5, and Y keys covered with red tape and all surrounding keys removed. In pre- and post-training sessions, participants were required to press the red keys with the five fingers of their left hand in a specified order. During the observational training, participants watched videos of the experimenter performing the keypress task. For the video recordings, a similar keyboard was used with the only difference that the sides of the five keys were covered in yellow to improve the visibility of the key being pressed. Stimuli presentations and response recordings were performed using MATLAB 8.3.0 (The MathWorks, MA, USA) and Psychophysics Toolbox 3.0.12 (Brainard, 1997).

Keypress sequences The same set of 12 five-element keypress sequences was used previously by Wiestler and Diedrichsen (2013). Each sequence required the five fingers of the left hand to press once, but in a different order and with no more than three adjacent finger-presses in a row. All sequences were matched for difficulty as demonstrated by prior work (Wiestler & Diedrichsen, 2013).

2.3.1. Videos

For observational training, 13-second videos were created showing the experimenter's left hand from a first-person perspective, slightly tilted to the right (Fig. 1B). Each video showed the experimenter executing one sequence five times, with naturally varying breaks between each sequence repetition, to ensure a more authentic presentation of the performance. That is, we tried to make the videos mimic natural motor performance as much as possible, so that the experience of physically practicing and observationally practicing were as similar as possible. For the same reason, for each sequence, five different video versions were recorded, to allow closer to natural performance variation of the same sequence. An additional video version for each sequence was created where one of the five sequence executions was incorrect. This resulted in 72 videos in total.

Sequences were executed at an intermediate baseline performance level, determined by behavioural pilot test results, where the average correct sequence execution at baseline was 2.29 s ($N = 17$, $M = 2.29$ s, $SE = 0.14$). Each original video, showing five repetitions of the same sequence, was slightly speeded up or slowed down ($\pm 10\%$) to make it exactly 13 s long. Consequently, some authenticity was lost; however, the relative variability within the video remained intact, and the average single sequence execution in the videos was 2.3 s. The videos were presented on a computer monitor in full colour on a black background. The frame rate was 29 frames per second with a resolution of 600×526 pixels, showing approximately natural hand size.

2.3.2. Sequence execution trial

A sequence execution trial involved five continuous repetitions of the same sequence (Fig. 1A). Participants were instructed to execute sequences as quickly and as accurately as possible. Each trial started with a 5-digit cue (for 2.7 s), indicating the sequence of keypresses. All trial-related information was presented centrally at the bottom of the screen against a grey background. A trial started with a black fixation cross (0.2 s), followed by the sequence cue presented as five digits (2.7 s) that indicated from right to left which key to press: "1" – the right-most key pressed with the thumb; "5" – the left-most key pressed with the little finger. After the cue, the digits were replaced by the fixation cross and five black asterisks above it. This served as a "go" signal to execute the memorised sequence five times as quickly and accurately as possible. If the correct key was pressed, the corresponding asterisk on the screen turned green, if a wrong key was pressed, the asterisk turned red.

After executing a single sequence, the central fixation cross changed colour giving feedback on the performance (0.8 s): green – correct sequence execution; red – incorrect sequence execution; blue – correct, but executed 20% slower than the median execution time in the previous trials; three green asterisks – correct and executed 20% faster than the median execution time in the previous trials. We chose to

include feedback in this manner, in order to compare with the physical practice study that the current approach was based upon (Wiestler & Diedrichsen, 2013). After this short feedback, all asterisks turned black signalling the start of the next execution trial. After five executions of the same sequence, the trial ended, and the next sequence was cued.

2.3.3. Sequence observation trial

A sequence observation trial involved watching a video clip of an actor's left hand performing five continuous repetitions of the same sequence (Fig. 1B). A trial started with a 5-digit cue (for 2.6 s¹), indicating the sequence to be executed, followed by a video (13 s) showing five executions of the cued sequence. Participants were instructed that they would see videos of a hand executing finger-press sequences and that they should watch and learn the sequences because they will have to perform the sequences at the end of the experiment. Participants were also instructed to watch whether the hand executed the correct (cued) sequence all five times because, occasionally, they would be asked to verify whether any errors were made. After some of the trials, participants were asked whether there was an error in any of the five executions – the error question. Participants responded to the error question by pressing a 'b' key (marked red) for 'yes' and an 'm' key (marked blue) for 'no'. This task was included to ensure that participants paid attention to the videos. Participants were also informed that they will need to perform the watched sequences again at the end of the experiment.

2.4. Procedure

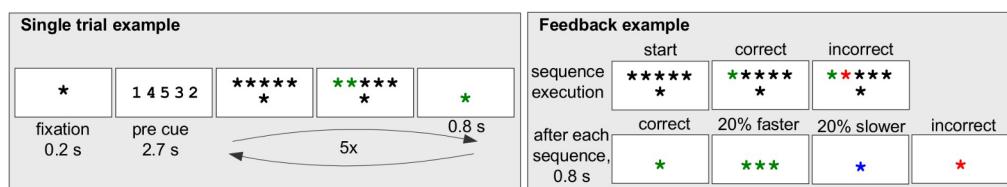
On arrival, participants were randomly assigned to physical (PP) or observational (OP) practice groups. For each participant, from the set of 12 sequences, one sequence was randomly allocated to aid familiarisation with the task, two other sequences to the Trained condition, and two more to the Untrained condition. The remaining sequences were unused. The task required learning two keypress sequences with the left (non-dominant) hand either by a physical practice (PP group) or by watching videos of an actor executing the sequences (OP group).

Familiarisation involved three single sequence execution trials to ensure participants understood the task. One trial consisted of five continuous repetitions of the same sequence. In the pre- and post-training sessions participants executed the two trained and two untrained sequence trials (one trial per sequence) in a random order. During training, participants practised two sequences by either performing (PP group) or watching (OP group) 36 trials of each sequence. The training session was divided into four sub-sessions. Each sub-session consisted of 9 trials per sequence. For the OP group, one of the 9 trials was an 'error trial' – a video showing at least one incorrect sequence execution. In each sub-session, the error question was asked randomly 5–7 times. The error question was included to assess the extent that participants were paying attention to the sequence information in the videos. Attention to the observed videos was assessed as a percentage of accurate responses to the error question.

The whole testing procedure lasted approximately two hours and consisted of the following steps: information, consent and instructions; Matrices test; motor task familiarisation; pre-test; 9 blocks of training; Big Five inventory; 9 blocks of training; IRI questionnaire; 9 blocks of training; NPI questionnaire; 9 blocks of training; Self-Construal scale questionnaire; post-test; Analogies test; Numbers test; spatial short-term memory test; debrief.

¹ The duration should have been modified to 2.7 s to match the duration during execution trials. 2.6 s was the duration used in an fMRI study using the same task (Apšvalka et al., 2018) and we failed to update the timing. We do not feel, however, that such a small discrepancy between conditions will make a meaningful difference to the overall pattern of results.

A. Sequence execution trial



B. Sequence observation trial



Fig. 1. Trial structures for sequence execution and sequence observation. (A) Sequence execution trial example. A cued sequence had to be memorised and then executed five times while receiving performance feedback. (B) Sequence observation trial example. A sequence cue was followed by a video showing a hand executing the sequence five times, either correctly or incorrectly. Occasionally a question was asked whether there was an error in any of the five repetitions, and a response had to be made.

2.5. Data analysis

All statistical analyses were performed using R (v3.3.2, 2016-10-31) in RStudio (v1.0.136, 2016-12-21, RStudio, Inc, Boston, MA). The significance threshold for all statistical comparisons was $p < 0.05$. All sample means are reported with their 95% confidence intervals in square brackets. Confidence intervals were calculated as $SE^*1.99$, for 91 degrees of freedom.

2.6. Measures of training effects on sequence learning

Participants' physical performance was assessed at pre- and post-test, measuring the average sequence execution time of the two trained (to-be-trained) and the two untrained sequences. Therefore, the average execution time was calculated across two sets of trials that each involved 5 repetitions of each sequence. The sequence execution time was measured as the duration between the first and the fifth keypresses. Incorrectly executed sequences were excluded from further analysis.

The effect of training on sequence-specific learning was assessed as a post-training percentage difference between the trained and untrained sequence execution times accounting for possible pre-training percentage differences between the sequences. The effect of training on sequence-specific learning was calculated according to Equation 1 below.

$$\text{Equation 1} = 100 * \left(\frac{\text{post}_{\text{Untrained}}\text{ET}}{\text{post}_{\text{Trained}}\text{ET}} - \frac{\text{pre}_{\text{Untrained}}\text{ET}}{\text{pre}_{\text{Trained}}\text{ET}} \right)$$

The effect of training on general learning was measured as pre to post percentage difference of the untrained sequence execution times according to Equation 2 below. We chose to measure general skill learning in this manner, in order to compare with the physical practice study that the current approach was based upon (Wiestler & Diedrichsen, 2013).

$$\text{Equation 2} = 100 * \left(\frac{\text{pre}_{\text{Untrained}}\text{ET}}{\text{post}_{\text{Untrained}}\text{ET}} - 1 \right)$$

Effect sizes for learning gains were reported in units of % increase as well as in standardised form as Cohen's d or dz , depending on whether the effect is a difference between groups (i.e., PP vs. OP) or a difference between conditions within the same participants (i.e., trained vs. untrained) (Lakens, 2013).

2.7. Data processing

One participant who reported being left-handed and eighteen participants who did not correctly execute any trials in one (or more) of the four conditions (pre-Trained, pre-Untrained, post-Trained, post-Untrained) were excluded from the analysis. More specifically, OP participants were excluded only due to pre-training errors. Two PP participants were excluded because at post-training they did not execute any of the untrained sequences correctly. One PP participant at post-training didn't execute any of the trained sequences correctly and one PP participant at post-test didn't execute any of the sequences correctly. Other excluded PP participants at pre-test didn't execute any of the to-be-trained or remain-untrained sequences correctly.

Additionally, twelve participants from the OP group were excluded due to more than a 50% error rate in response to the 'error question' in the second, third or fourth training sub-session. The exclusion was based on the assumption that the first sub-session entailed familiarisation with the task and having more than 50% error rate on the following sub-sessions would indicate a lack of attention to the observed videos, thus potentially compromising a practice effect.

From the remaining sample, eight participants were excluded as pre-test outliers. The outliers were defined as pre-Trained or pre-Untrained execution time values that were more than two times the interquartile range above the third quartile or below the first quartile. Of the remaining 184 participants, the mean-average incorrect trial removal rate across pre-test and post-test was 23.02% ($SD = 12.59$).

2.8. Hypothesis testing

Our main hypothesis was that fluid intelligence and working memory would predict sequence-specific training effects for both PP and OP groups. In addition, given the different processes that may underpin sequence-specific and general learning (Janacek & Nemeth, 2013; Wong, Lindquist, Haith, & Krakauer, 2015), fluid intelligence and working memory may have dissociable effects on general compared to sequence-specific learning. Therefore, we also tested the extent to which fluid intelligence and working memory made dissociable contributions to sequence-specific and general learning. We used multiple regression to test these hypotheses. PP and OP groups were analysed separately and all variables were converted to within-group z-scores. The benefit of scaling in this manner prior to analysis is that it makes it easier to interpret the data as it returns standardized beta coefficients. However, we also ran our primary analyses without converting the data

to z-scores and the results remained the same.

The regression models consisted of the training effect as the dependent measure and three predictor variables: baseline performance (an inverse of the pre-training average of trained and untrained sequence execution times; shorter execution time equates to higher performance), as well as fluid intelligence and working memory scores. The baseline performance was included as a predictor because participants who are already skilled at the task may have little benefit from the training compared to participants with poorer initial skills (Alexander & Smales, 1997).

In terms of sensitivity to detect these primary effects, given the sample size of 92 participants in each group, we had 80% power to detect effects that are conventionally considered small to medium ($f^2 = 0.12$; Cohen, 1988). Effect size sensitivity was estimated with a *pwr.f2.test* function in R for a linear regression model with three predictor variables and sample size 92.

In addition to these tests of our main hypotheses, we also performed an exploratory analysis, which investigated whether personality traits further explain the variance of the training effect (Supplementary Materials).

3. Results

Sequence execution times at pre-test, post-test and during practice sessions are illustrated in Fig. 2. Training gains across each group, as well as for individual participants, are illustrated in Fig. 3. In addition, we use scatter plots to illustrate the raw data for relationships across participants between learning rates and our key individual differences measures (Fig. 4). Key statistics from our regression models are reported in Tables 2 and 3 and visualised in Fig. 5.

3.1. Group characteristics

The PP and OP groups were compared using a Chi-square test on the proportion of males and females as well as the number of native English speakers. Participants' baseline performance, working memory and fluid intelligence scores were compared using an independent measures t-test. Personality questionnaire scores were compared using Mann-

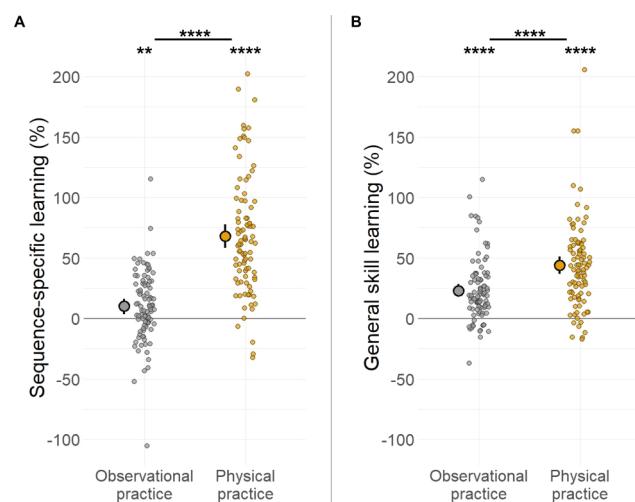


Fig. 3. Training effects on sequence-specific (A) and general skill (B) learning for the observational practice (grey) and physical practice (orange) groups. Large dots: group averages and 95% CI; small dots: individual participant values; ** $p < 0.01$, **** $p < 0.0001$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Whitney U tests. There were no significant differences between the two groups in terms of demographics, baseline performance or personality measures (Table 1).

3.2. Training effects on sequence learning

3.2.1. Sequence-specific learning

Both PP ($M = 68\%$ [58%, 78%], $t_{91} = 13.44$, $p < 0.0001$, $d_z = 1.40$) and OP ($M = 10\%$ [4%, 16%], $t_{91} = 3.32$, $p = 0.0013$, $d_z = 0.35$) groups showed significant training effects on sequence-specific learning (Fig. 3A). According to Cohen's benchmark criteria for interpreting effect sizes (Cohen, 1992), the effect size for the PP group is conventionally considered large or very large, whereas the effect size

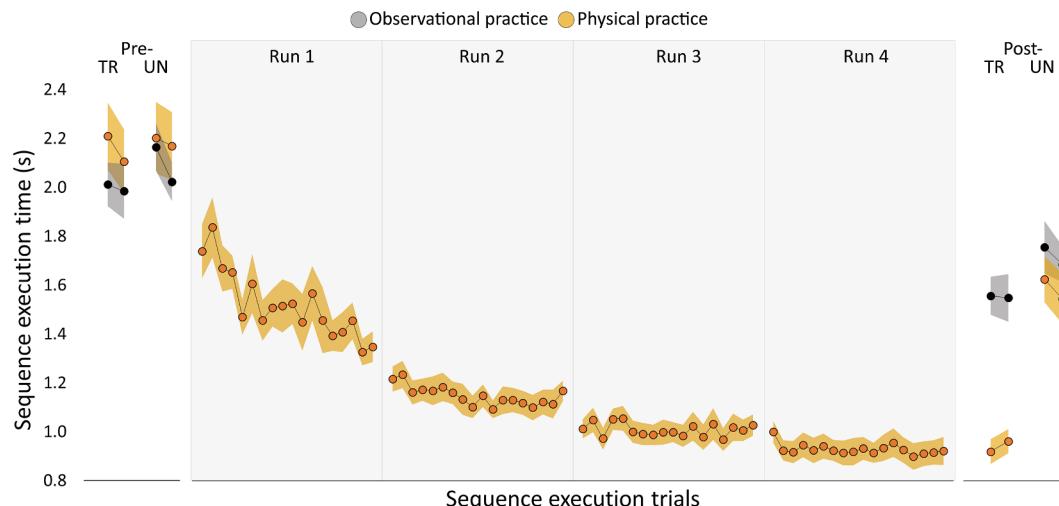


Fig. 2. Sequence execution time at pre-test, post-test and during practice sessions across observational and physical practice groups. As two sequences were allocated to each condition (trained and untrained), two data points are plotted per condition at pre-test and post-test. Similarly, during practice sessions, the physical practice group practised two sequences (note: the observational practice group did not physically practice, which is why no training data are reported for that group during training). Practice was divided into four sub-sessions (displayed as Run 1–4). Each sub-session consisted of nine trials per sequence (which produced 18 trials per sub-session in total) and a single trial comprised five consecutive executions of a sequence. Sequence execution time during practice was measured as an average of correct sequence executions within the trial. As such, if all five executions were performed incorrectly, the trial was not included in the plot. Therefore, the number of participants who contribute to each trial varies slightly from trial to trial (range 88–92; from a total of 92 participants in the physical practice group). Error bars represent within-participant 95% confidence intervals. Abbreviations: Tr. = trained; Untr. = untrained.

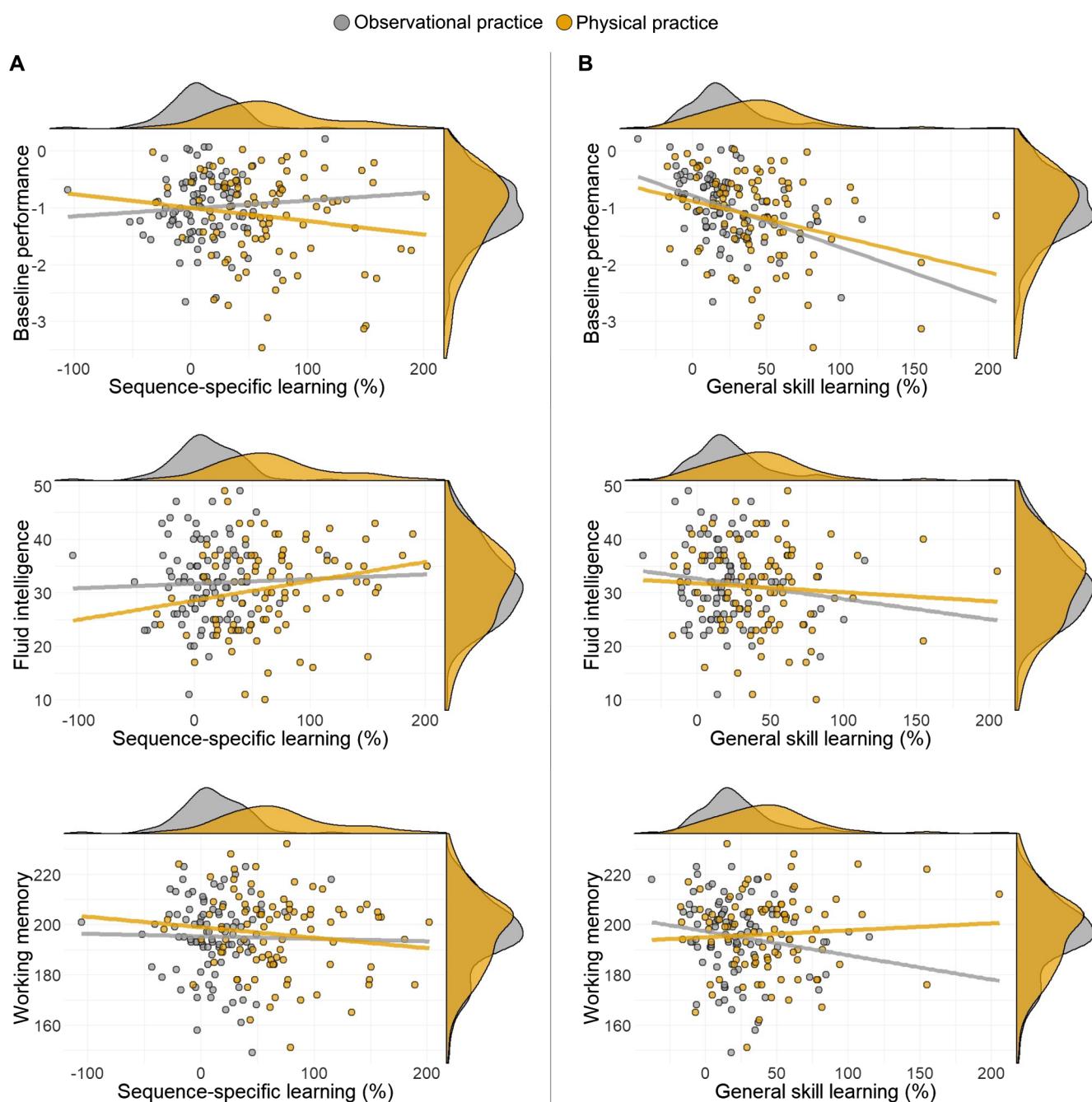


Fig. 4. Scatter plots and marginal distribution densities showing sequence-specific (A) and general skill (B) learning versus baseline performance, fluid intelligence and working memory for observational practice (grey) and physical practice (orange) groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Regression analysis summary of sequence-specific learning.

Model	Physical practice			Observational practice		
	F _{3,88}	p	R ²	F _{3,88}	p	R ²
Coefficients	β [95% CI]	t	p	β [95% CI]	t	p
Intercept	0	0	1	0	0	1
Baseline performance	-0.203 [-0.456, 0.050]	-1.601	0.112	0.076 [-0.161, 0.313]	0.637	0.526
Fluid intelligence	0.373 [0.149 0.597]	3.312	0.001	0.026 [-0.227, 0.278]	0.202	0.841
Working memory	-0.165 [-0.419, 0.089]	-1.291	0.200	-0.053 [-0.292, 0.185]	-0.446	0.657

Table 3
Regression analysis summary of general skill learning.

Model	Physical practice			Observational practice		
	F _{3,88} = 6.386, p = 0.0006, R ² = 0.179			F _{3,88} = 6.582, p = 0.0005, R ² = 0.183		
Coefficients	β [95% CI]	t	p	β [95% CI]	t	p
Intercept	0	0	1	0	0	1
Baseline perf.	-0.514 [-0.760, -0.268]	-4.158	0.00007	-0.434 [-0.650, -0.219]	-4.02	0.0001
Fluid intelligence	-0.268 [-0.245, 0.191]	-0.245	0.807	0.082 [-0.147, 0.311]	0.715	0.476
Working memory	0.384 [0.137, 0.631]	3.089	0.003	-0.072 [-0.288, 0.144]	-0.664	0.508

for the OP group was small to medium in size. Moreover, the training effect was considerably larger for the PP than the OP group ($M = 58\%$ [46%, 70%], $t_{149.31} = 9.80$, $p < 0.0001$, $d = 1.60$).

3.2.2. General learning

A post-training improvement of performance on untrained sequences, which reflects general skill learning, was expected at the group-level based on prior research using these sequence-learning designs (Janacek & Nemeth, 2013; Meier & Cock, 2014). Both PP ($M = 44\%$ [36%, 61%], $t_{91} = 11.41$, $p < 0.0001$, $d_z = 1.19$) and OP ($M = 23\%$ [18%, 29%], $t_{91} = 8.35$, $p = 0.0013$, $d_z = 0.87$) groups showed significant effects on general skill learning (Fig. 3B). According to Cohen's benchmark criteria for interpreting effect sizes (Cohen, 1992), the effect size for both groups is conventionally considered large or very large. The training effect was larger for the PP than the OP group ($M = 21\%$ [30%, 11%], $t_{165.33} = 4.38$, $p < 0.0001$, $d = 0.68$).

The way we chose to measure general skill learning, which was based on prior work using a similar paradigm (Wiestler & Diedrichsen, 2013), means that although the overall sequences that were trained differed from those that were untrained, there could be subcomponent transitions that are shared. For example, pairs or triplets of finger presses could be common between some trained and untrained sequences. It is possible, therefore, that training on some of these subsequences could spill over into the untrained performance measures and contribute in part to performance gains in general learning.

However, we are not concerned about this possibility for two reasons. First, overlap is relatively small, on average. The number of overlapping triplets between trained and untrained sequences was:

$M = 4\%$, $SD = 11\%$, range [0%, 50%]. Furthermore, a minority of participants – 29 participants out of 184 (16%) – had at least one overlapping triplet. The number of overlapping doubles between trained and untrained sequences was: $M = 28\%$, $SD = 17\%$, range [0%, 75%]. In this case, a majority of participants – 168 participants out of 184 (91%) – had at least one overlapping double. Second, if we remove participants with at least one overlapping triplet or 50% or more overlapping doubles, the results from our analyses of general skill learning remain unchanged. This includes pre to post differences in performance, as well as individual difference analyses. Therefore, our primary findings regarding general skill learning remain unchanged if only participants with a minority of shared transitions are included.

3.2.3. Exploratory observation: Negative learning

Although on average, across both practice groups, practice did lead to improvements in behavioural performance, we note that there are a minority of individuals that show negative learning indexes (i.e., performance deteriorates with practice; see Fig. 3). Such findings are not straightforward to interpret because most prior work only reports summary statistics for the group as a whole. Indeed, a strength of the current analysis is that the observation of individuals with negative learning profiles was made possible by plotting a data point for each participant in addition to summary statistics for the group as a whole.

Since we do not know how to interpret negative learning or explain why a minority show it, to aid future research here we report the proportion of individuals showing negative learning and offer some speculative considerations. Of the 92 OP participants, 31 (34%) exhibited negative sequence-specific learning, and 15 (16%) exhibited

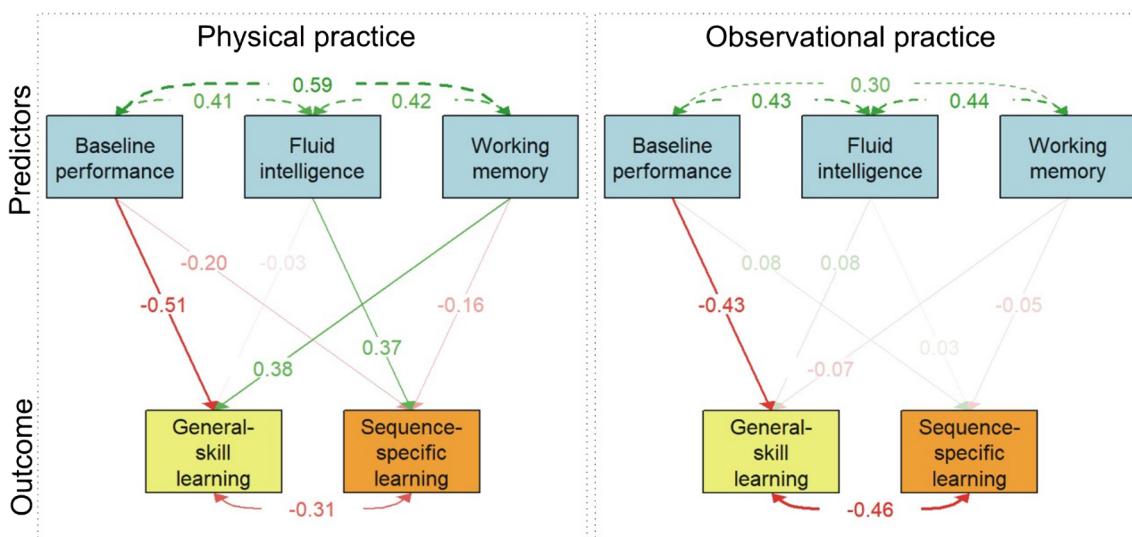


Fig. 5. Illustration of regression analyses for physical and observational practice groups. For each group (physical practice and observational practice), the figure shows standardised beta estimates of how baseline performance, fluid intelligence and working memory predict the sequence-specific and general skill learning. In addition, the figure shows the predictor intercorrelation coefficients and outcome intercorrelation coefficients. Green: positive, red: negative, fading reflects significance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

negative general skill learning. No participants exhibited both types of negative learning, which is consistent with the negative relationship that we report in [Section 3.3.2](#) between sequence-specific and general learning rates. Of the 92 PP participants, 4 (4%) exhibited negative sequence-specific learning, and 7 (8%) exhibited negative general skill learning. We suggest that for the PP group, this is rather negligible. But in the OP group, it is larger and warrants further discussion. It is possible that the observational practice participants in our study were not sufficiently challenged by the comparatively slow model performer they observed during practice. Indeed, fast performers may have learned from the slow model to slow down their performance – that is, they learned via observation, but learned to match the model rather than learning to get faster than the model. To support this speculation, most of the negative learners at pre-test performed better than the model. Specifically, in the OP group, 67 of the 92 participants at pre-test performed better than the model on average (avg. ET < 2.29 s). Of those 67, 27 exhibited negative sequence-specific learning and 14 exhibited negative general learning. Again, we offer caution when interpreting negative learning and hope that future research can probe what might be an interesting avenue for individual difference research to pursue.

3.3. Fluid intelligence and working memory as predictors of learning

3.3.1. Sequence-specific learning

We used multiple regression analysis to test whether fluid intelligence and working memory predict the sequence-specific training effect. The baseline performance (an inverse of the pre-training average of trained and untrained sequence execution times) was also included in the model to control for the baseline performance differences, which may contribute to the training effect. All three predictor variables were intercorrelated but not so highly as to suggest multicollinearity (e.g., $r > 0.80$). Fluid intelligence and working memory were positively correlated ($r = 0.432$, $p < 0.001$), and both fluid intelligence ($r = 0.423$, $p < 0.001$) and working memory ($r = 0.465$, $p < 0.001$) were positively correlated with the baseline performance.

The model that included the three predictor variables significantly explained sequence-specific training effect variance in the PP group ([Table 2](#); [Fig. 5](#)). However, fluid intelligence was the only individually significant predictor. When controlling for baseline performance and working memory capacity, fluid intelligence explained 14% of the training effect variance, such that higher fluid intelligence was associated with greater training gains ([Table 2](#); [Fig. 5](#)). The effect sizes for baseline performance (4%) and working memory (3%) were much smaller and did not pass our threshold for statistical significance. [Fig. 4A](#) illustrates the raw data in scatter plots.

To check that the component correlation pairs in our model were not unduly driven by outliers in the data, we also performed robust correlations using the robust correlation toolbox (Pernet, Wilcox, & Rousselet, 2013). To do so, we ran Pearson Skipped correlations between our three predictor variables and the results remained the same, such that they were all positively correlated. In addition, we computed the same Pearson Skipped correlation between fluid intelligence and sequence-specific learning, and this also remained positively correlated ($r = 0.23$, 95%CI [0.05, 0.40]). Therefore, these results suggest that the component correlations within our model are robust in the sense that they are not unduly influenced by outliers.

Contrary to our predictions, none of our three key predictor variables explained variance in sequence-specific training effects in the OP group ([Table 2](#); [Fig. 5](#)). Moreover, in terms of effect sizes and interval estimates, all three predictors explained less than 1% of the variance in sequence-specific training gains and their 95% confidences intervals were squarely overlapping with zero, which is suggestive of no effect. Thus, as illustrated in [Fig. 5](#), fluid intelligence and working memory were not associated with a higher sequence-specific training effect following OP. [Fig. 4A](#) illustrates the raw data in scatter plots.

3.3.2. General skill learning

The three-predictor model significantly explained general skill learning variance in both PP and OP groups ([Table 3](#); [Fig. 5](#)). In the PP group, lower baseline performance and higher working memory significantly predicted a higher training effect on general skill learning (26% and 15% respectively). Pearson Skipped correlations showed that the relationship between baseline performance and general skill learning was significant and robust ($r = -0.24$, 95%CI [-0.42, -0.03]), but the relationship between working memory and general skill learning was not significant and less robust ($r = -0.11$, 95%CI [-0.31, 0.092]). Of course, the zero-order correlation between working memory and general skill learning does not include baseline performance as it does in the model presented above. By contrast, fluid intelligence predicted a smaller proportion of general skill learning (7%). Further, fluid intelligence was not a significant predictor of general skill learning and the confidence intervals squarely overlapped with zero.

Also, in the OP group, lower baseline performance (19%) predicted higher general skill learning ([Table 3](#); [Fig. 5](#)). Neither fluid intelligence nor working memory were significant predictors of general skill learning following OP with each predictor explaining less than 1% of the variance and 95% confidence intervals overlapping with zero. [Fig. 4B](#) illustrates the relevant raw data in scatter plots.

An integrated visualisation of relationships among the involved measures is presented in [Fig. 5](#). For each group, in addition to the standardised beta estimates of the two regression models, the figure shows positive correlations among the three predictor variables and a negative correlation between the general skill learning and sequence-specific learning. Overall, fluid intelligence and working memory were significant predictors of different components of the physical practice effects, but none of the variables predicted observational practice effects. Of course, it is worth acknowledging that under a null hypothesis significance testing statistical framework, failure to reject the null hypothesis is not the same as supporting the null hypothesis. Therefore, whilst our best current estimate based on point and interval estimates is that observational learning rates do not vary as a function of our key predictors, we cannot provide formal support for the null. Instead, based on our power analysis, we can be reasonably confident that should an effect at least as large as $f^2 = 0.12$ exist, we would have been able to detect it in the current study. As such, we can conclude that any effect of working memory or fluid intelligence on observational learning is likely to be near zero or negligible in size.

3.4. Fluid intelligence and working memory as predictors of task-focussed learning

It is important to emphasise that during practice sessions the OP group had two parallel tasks: learn the motor sequence and detect errors in the observed model's performance. However, during observational practice sessions, an actual response was only required in reference to the error detection task. As such, we were interested to see whether fluid intelligence and working memory are related to the improvements in error detection, rather than sequence learning. If so, it may suggest that participants prioritised learning strategies aimed at detecting errors more than learning sequences.

Across the four observational practice sub-sessions (runs), the mean error detection accuracy in the OP group was 89% [87%, 91%]. There was a significant improvement from run 1 to run 2 ($t_{91} = 3.99$, $p = 0.0001$) with no significant improvements in the following runs ($p > 0.380$; [Fig. 6A](#)). We excluded run 1 from the subsequent analysis assuming that during the first sub-session, error detection accuracy reflected a familiarisation phase. Therefore, observational practice-related error detection improvement was measured as error detection accuracy difference between run 2 and run 4 (however, the results of run 1 vs. run 4 also show the same pattern). Although on average there was no significant difference in error detection between runs two and

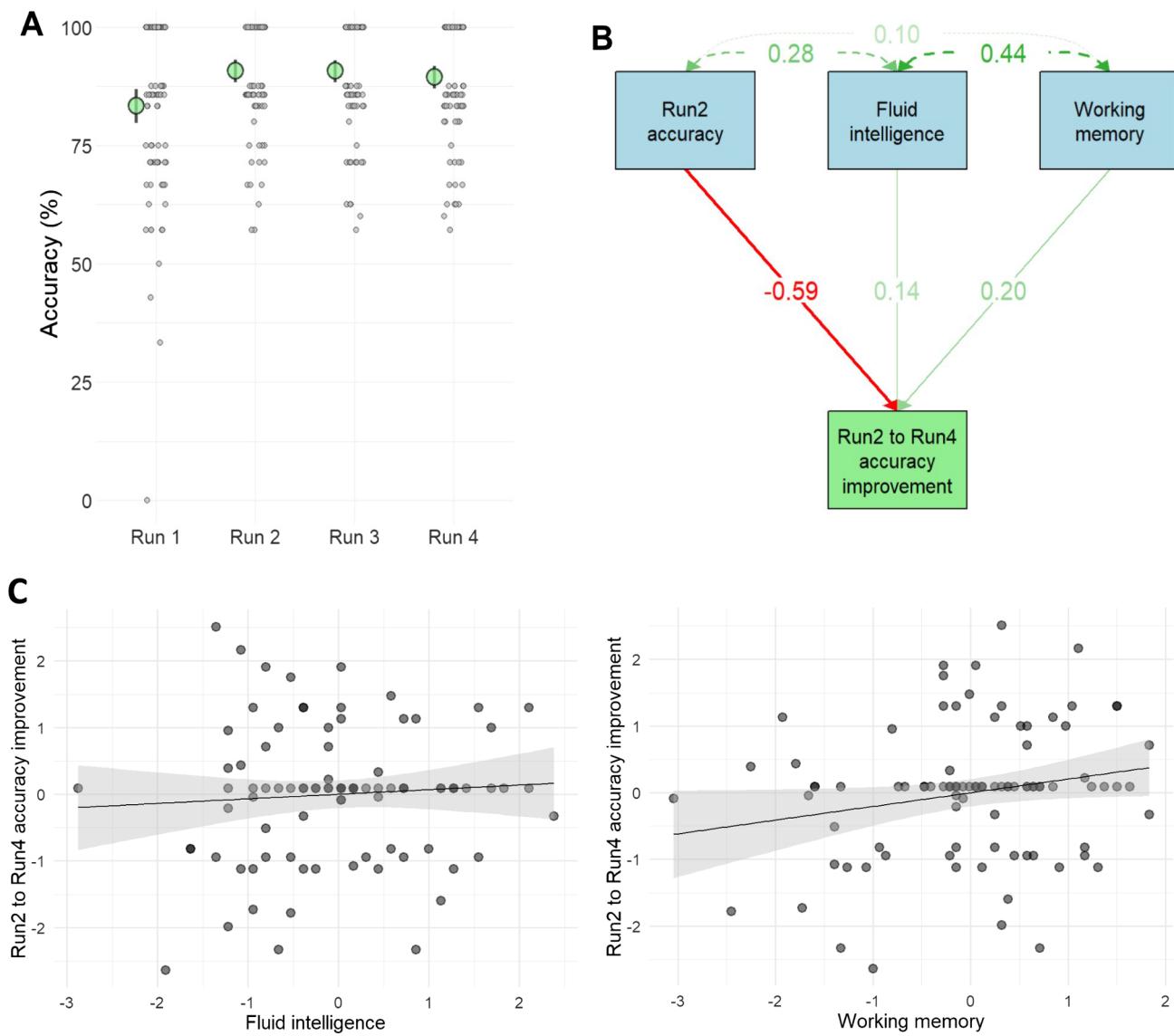


Fig. 6. Error detection accuracy and perceptual improvement predictors. (A) Group-averaged accuracy in response to the error question during observational training. Large dots: group averages and 95% CI; small dots: individual participant values. (B) Perceptual improvement predictor variables. The figure shows standardised beta estimates of how error detection accuracy in run 2, fluid intelligence, and working memory predict the perceptual improvement from run 2 to run 4. In addition, the figure shows the predictor intercorrelation coefficients. Green: positive, red: negative, fading reflects significance. (C) Scatter plots showing the raw data for changes in accuracy improvement from run 2 to run 4 as a function of fluid intelligence (left panel) as well as working memory (right panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

four, we wanted to investigate individual differences in participants' error detection improvements. Importantly, we were able to measure only the general (pre- to post- training improvement) and not the sequence-specific improvement as participants were never asked to watch the untrained sequences during training.

The error detection accuracy at run 2 (as a baseline performance), fluid intelligence, and working memory measures were z-scored and included in a multiple regression analysis to test whether they predict the error detection accuracy improvement from run 2 to run 4. The regression model significantly explained 11% variance in error detection, with lower baseline performance (35%) and higher working memory (4%) as significant predictors (Table 4; Fig. 6B). To show the raw data, we use scatter plots to visualise the relationship between change in accuracy score and fluid intelligence, as well as working memory (Fig. 6C). Consistent with the modelling results, Pearson Skipped correlations show that working memory is modestly associated with changes in accuracy detection $r = 0.21$ 95%CI = [0.00, 0.40], whereas fluid intelligence explains very little of the variance $r = 0.01$

Table 4

Observational practice group error detection improvement (from run 2 to run 4) regression analysis summary.

Model	Run4 – Run2 accuracy		
	F _{3,88} = 16.56, p < 0.0001, R ² = 0.334		
Coefficients	β [95% CI]	t	p
Intercept	0	0	1
Run2 accuracy	-0.588 [-0.765, -0.411]	-6.618	< 0.0001
Fluid intelligence	0.144 [-0.052, 0.341]	1.463	0.147
Working memory	0.203 [0.014, 0.393]	2.133	0.036

95%CI = [-0.19, 0.21]. A similar set of results was observed for the general motor skill learning in the PP group (Table 3). Overall, in the OP group, working memory was a significant predictor for improvements in the error detection task, but not for the motor skill learning task.

4. Discussion

Although evidence suggests that learning by physical and observational practice relies on partly shared cognitive and neural systems, no research to date has investigated the extent to which common individual differences predict both types of learning. In the current study, by taking an individual differences approach to study skill acquisition, we provide a novel test of the extent to which common cognitive systems underpin learning following observational and physical practice. Here we show that individual differences in working memory and fluid intelligence predict improvements in dissociable aspects of motor learning following physical practice, but not observational practice. Therefore, these results suggest limits to the extent that learning through different forms of experiences, such as physical and observational practice, rely on shared mechanisms. Consequently, models of observational learning need updating to reflect the extent to which such learning is based on shared as well as distinct processes compared to physical learning. We suggest that such differences could reflect the more intentional nature of learning during physical compared to observational practice, which relies to a greater extent on higher-order cognitive resources such as working memory and fluid intelligence. One potential implication of these findings is that lower-order sequence-learning processes, which are associated with visuomotor integration and motor planning, may be more invariant across individuals. More broadly, these findings have implications for learning interventions that aim to improve skill acquisition, as it may point towards the types of mechanisms involved in different modes of learning, as well as the type of cognitive systems that may be more or less amenable to modification. In the following, we discuss several implications that the findings have for understanding cognitive mechanisms of physical and observational practice, as well as some of the limitations of the research, which should also help to guide the interpretation of the findings.

4.1. Individual differences in skill learning through physical and observational practice

Our findings demonstrate that working memory and fluid intelligence predict dissociable aspects of learning through physical practice. Fluid intelligence predicts sequence-specific motor learning beyond the influence of working memory. In contrast, working memory predicts general learning improvements across all sequences from pre-to post-training (Fig. 7). Therefore, although working memory and fluid intelligence are correlated, they also support different processes in learning (Ackerman et al., 2005; Kane et al., 2005; Shipstead et al., 2016; Wang et al., 2017). Indeed, the results mirror findings observed in non-motor contexts, such as when learning abstract rules, where fluid intelligence has been shown to predict learning and retrieval processes above and beyond working memory (Wang et al., 2017).

It has been suggested previously that working memory might relate more to general skill learning rather than sequence-specific learning (Janacek & Nemeth, 2013; Rhodes, Bullock, Verwey, Averbeck, & Page, 2004). Working memory is important in supporting attention and maintaining task goals (Unsworth & Engle, 2005). These abilities are essential for general task performance, which relies on short-term

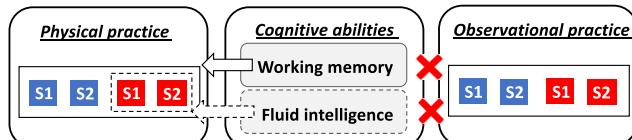


Fig. 7. A schematic illustration of the relationships between cognitive abilities and skill learning through physical and observational practice. Blue = untrained. Red = trained. S = Sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

memorisation of the cued sequence and fast execution of discrete keypresses. By contrast, sequence-specific skills additionally involve long-term memory retrieval of the trained sequence and integration of its discrete keypress elements into a unified sequence representation (Abrahamse, Ruitenberg, de Kleine, & Verwey, 2013; Verwey, 1996). Moreover, selective retrieval of relevant information from long-term memory is a crucial component of fluid intelligence (Unsworth & Engle, 2005). Therefore, these results demonstrate that fluid intelligence predicts a measure of learning that better reflects processes that are specifically tied to learning action sequences, rather than improvements in task performance more generally (Janacek & Nemeth, 2013; Wong et al., 2015).

Similar to learning through physical practice, learning through observation involves high-order cognitive processes (Hodges, Ong, Larsen, & Lim, 2011; Lim, Larsen, & Hodges, 2014; Maslovat, Hodges, Krigolson, & Handy, 2010; Vogt & Thomaschke, 2007). Although many similarities exist between the cognitive and neural processes supporting learning through physical and observational practice (Blandin et al., 1999; Boutin et al., 2010; Gardner et al., 2017; Hodges et al., 2007; Vogt & Thomaschke, 2007), here we show limits to this similarity. Indeed, as tested here, working memory and fluid intelligence predict dissociable aspects of learning in physical, but not observational practice (Fig. 6). Thus, performance gains following observational practice seem largely indifferent to variation in working memory and fluid intelligence across individuals.

The lack of variation in performance gains according to our key individual differences following observational learning warrants further consideration from a statistical viewpoint. Within a null hypothesis significance testing statistical framework, it is important to remember that failure to reject the null hypothesis is not the same as supporting the null hypothesis. Indeed, to help guide the interpretation of the results, any failure to reject the null hypothesis must be qualified by the sensitivity of the design. If we consider the sensitivity of our primary measure, therefore, we had 80% power to detect small to medium effects ($f^2 = 0.12$), which means we can be relatively confident that if an effect of this magnitude or higher did exist, we would have been able to detect it. As a consequence, our best estimate of the influence of working memory and fluid intelligence on learning through observational practice, is that if an effect does exist, it is likely to be close to zero or relatively small in magnitude (i.e., smaller than $f^2 = 0.12$). At a minimum, therefore, our results demonstrate that, as tested here, working memory and fluid intelligence play distinct roles in learning through physical and observational practice. Moreover, for future research to convincingly test for the presence of smaller effects (i.e., less than $f^2 = 0.12$), considerably larger sample sizes would be required.

There are many ways to learn through observing others and we expect that there are circumstances where cognitive abilities would indeed predict learning rates through observation. As such, it is important to consider constraints on the generality of our findings (Simons, Shoda, & Lindsay, 2017). For example, situations that place more demand on intentional learning strategies during the observation of others may reveal individual differences in learning that resemble more closely those observed during physical practice. Support for this suggestion is provided by previous research, which shows that fluid intelligence and working memory are significant predictors for intentional, but not unintentional, learning. Under intentional learning conditions, individuals engage in heightened cognitive control processes including the regulation of attention and executive control (Maxwell, Masters, & Eves, 2003; Norman, Price, & Duff, 2006; Unsworth & Engle, 2005). Such differences in learning strategies and associated cognitive processes between physical and observational learning may underlie different patterns of individual differences.

Findings from the current study also support the proposed relationship between the intentionality of sequence learning and individual differences. The observational practice group was given explicit instructions to perform two tasks: learn sequences and detect

errors. However, only the error detection task was monitored during training sessions, which may have placed more focus and attentional resources on detecting errors than learning sequences. In support of this proposal, following observational practice, we found that working memory was a significant predictor of the improvement in error detection accuracy, but not for the improvement in keypress sequence performance. This finding mirrored the relationship between working memory and pre- to post-test improvements in sequence learning through physical practice. As such, we suggest that sequence-learning by observation became more of a secondary task compared to error detection, thus possibly explaining why cognitive abilities did not emerge as reliable predictors of practice effects. Indeed, previous research shows that implicit learning has little variation across individuals (Kaufman et al., 2010; Reber, Walkenfeld, & Hernstadt, 1991), whereas explicit or intentional learning varies as a function of cognitive abilities (Christou, Miall, McNab, & Galea, 2016).

Although we acknowledge that including the error detection task may have altered the attentional focus away from observational learning, we chose to include it in order to be able to identify how engaged participants were in the process of learning through observation. If we had chosen to leave out such a task, we would have no assurance that participants were actually observing the videos during practice, which was crucial to our study. As such, we were left with an experimental trade-off with no obviously superior approach. Moreover, although this design choice could be one reason for the different results between the two training groups in terms of working memory, it is unlikely to account for the results regarding fluid intelligence. Indeed, it seems reasonable to argue that attentional focus and consequent deployment of working memory resources could be altered by the choice of task during physical compared to observational learning. However, it is difficult to argue how fluid intelligence could be affected in the same way by the choice of task. Therefore, we suggest more caution should be taken when interpreting the relationship between working memory and observational learning than fluid intelligence and observational learning. In addition, a further feature of the experimental procedure in our study may have contributed to the unintentional nature of sequence learning in the observational practice group. The main task for the physical practice group was fast and accurate execution of the cued sequences, receiving constant feedback, thus encouraging performance improvement. In retrospect, we acknowledge that physical practice without feedback would have been more appropriate for comparing the effects of learning by physical and observational practice (Kirsch & Cross, 2015).

Although differences in learning focus (action sequences vs. error detection) may contribute to the different results observed in terms of individual differences, these results have important implications for ecological considerations regarding observational learning. That is, in many social contexts, learning through observation will frequently be task-independent (c.f., Cross et al., 2009). Of course, there are times when we may want to observe and intentionally learn a new skill, such as how to perform a pirouette in ballet or how to change a flat tyre. As such, it will be valuable for future studies to examine the extent to which individual differences in cognitive abilities influence learning by observation compared to physical practice across a much broader range of task complexities, such as learning to juggle (Hodges & Coppola, 2015), tie knots (Cross, Hamilton, Cohen, & Grafton, 2017), play the guitar (Gardner et al., 2017) or dance (Kirsch & Cross, 2015). But in many social instances, we will be engaged in a primary task (make a cup of tea or chat with friends) and passively learn regularities in the environment from watching others. In contrast, in physical practice, we are more likely to have a pre-determined intention to learn and therefore practice a new task. Hence, the operation of cognitive mechanisms that underpin observational learning, which is typical in everyday life, may be relatively invariant across individuals. These relatively invariant cognitive mechanisms may reflect relatively lower-order sequence-learning processes, which are associated with visuomotor integration and motor planning.

4.2. Broader implications for models of physical and observational practice

Although common cognitive and neural processes are implicated in action perception and production, as well as during physical and observational learning, our results highlight the importance of testing limits to such shared mechanisms. As such, models of observational learning should account for common cognitive processes, which are likely to be shared with learning by physical practice, as well as distinct processes that might also be at play. Moreover, these mechanisms may be flexibly deployed depending on the learning context, such that there may be different sub-types of observational learning (some more intentional and others more passive, for example). A different combination of processes may be involved depending on such contexts.

Relatedly, models of observational learning that explicitly distinguish between action-specific (i.e., domain-specific) processes and domain-general processes, such as the regulation of attention and executive control, would be valuable. Such models have emerged in the domain of sequence execution (Abrahamse et al., 2013; Verwey, 2001; Verwey, Shea, & Wright, 2015) and these may offer fertile ground to build upon for models of observational learning. This is particularly pertinent given the wide variety of frontoparietal brain regions that have been implicated in observational learning (Cross et al., 2009; Higuchi et al., 2012; Kirsch & Cross, 2015; Sakreida et al., 2018; Vogt & Thomaschke, 2007). Many of these frontoparietal regions overlap with key nodes in action representation systems, such as the mirror neuron system (Rizzolatti & Sinigaglia, 2016), as well as domain-general control processes that have been associated with the multiple demand network (Duncan, 2010). Therefore, studies that distinguish between different functional contributions of frontoparietal cortex would be valuable.

Furthermore, models that distinguish between domain-specific and domain-general contributions could also take inspiration from models of semantic and social cognition that stress the interplay between domain-specific representational content and domain-general ‘control’ of such representations (Barrett, 2012; Binney and Ramsey, 2019; Jefferies, 2013; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017; Michael & D'Ausilio, 2015; Ramsey, 2018; Spunt & Adolphs, 2017). Indeed, in physical practice, for example, it could be the interplay between representation and control that may vary as a function of individual differences in cognitive abilities. Moreover, it is this interplay between cognitive components that may differ depending on working memory and fluid intelligence – that is, fluid intelligence is likely to predict tighter links between control systems and task-specific representations. In contrast, this interplay between representational content and control is likely to be minimised in relatively passive observational learning, as cognitive resources are deployed to other task features. The broader point is that models need to be specified before they can be tested (Gray, 2017), which means results should generate hypotheses that can be tested in future studies. Moreover, this approach is consistent with placing a greater emphasis on central and perceptual processes when attempting to understand motor performance (Rosenbaum, 2005; Rosenbaum, Chapman, Coelho, Gong, & Studenka, 2013).

Acknowledgments

This work was supported by the Ministry of Defence of the United Kingdom Defence Science and Technology Laboratory [grant number DSTLX-1000083177 to ESC and RR], the Economic and Social Research Council [grant numbers ES/K001884/1 to RR and ES/K001892/1 to ESC], a Marie Curie Actions/FP7 [CIG11-2012-322256 to ESC], and a European Research Council grant [ERC-2015-StG-677270 to ESC].

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2019.04.015>.

- personality inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, 54(5), 890–902.
- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 888–896. <https://doi.org/10.1037/0278-7393.17.5.888>.
- Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. P. a. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science*, 23(5), 699–746. <https://doi.org/10.1016/j.humov.2004.10.008>.
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4), 264–274. <https://doi.org/10.1038/nrn2805>.
- Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience*, 17(12), 757.
- Rosenbaum, D. A. (2005). The Cinderella of psychology: The neglect of motor control in the science of mental life and behavior. *American Psychologist*, 60(4), 308.
- Rosenbaum, D. A., Chapman, K. M., Coelho, C. J., Gong, L., & Studenka, B. E. (2013). Choosing actions. *Frontiers in Psychology*, 4, 273.
- Sakreida, K., Higuchi, S., Di Dio, C., Ziessler, M., Turgeon, M., Roberts, N., & Vogt, S. (2018). Cognitive control structures in the imitation learning of spatial sequences and rhythms—An fMRI study. *Cerebral Cortex*, 28(3), 907–923. <https://doi.org/10.1093/cercor/bhw414>.
- Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, 11(6), 771–799. <https://doi.org/10.1177/1745691616650647>.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>.
- Simmons, Joseph P., Nelson, Leif D., & Simonsohn, Uri, A 21 Word Solution (October 14, 2012). Available at SSRN: <https://ssrn.com/abstract=2160588>.
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1177/1745691617708630>.
- Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, 20(5), 580–591. <https://doi.org/10.1177/0146167294205014>.
- Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: Insights from social neuroscience. *Nature Reviews Neuroscience*, 18(9), 559–567. <https://doi.org/10.1038/nrn.2017.76>.
- Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition*, 33(2), 213–220. <https://doi.org/10.3758/BF03195310>.
- Verwey, W. B. (1996). Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3), 544–562. <https://doi.org/10.1037/0096-1523.22.3.544>.
- Verwey, W. B. (2001). Concatenating familiar movement sequences: The versatile cognitive processor. *Acta Psychologica*, 106(1–2), 69–95.
- Verwey, W. B., Shea, C. H., & Wright, D. L. (2015). A cognitive framework for explaining serial processing and sequence execution strategies. *Psychonomic Bulletin & Review*, 22(1), 54–77. <https://doi.org/10.3758/s13423-014-0773-4>.
- Vogt, S., & Thomaschke, R. (2007). From visuo-motor interactions to imitation learning: Behavioural and brain imaging studies. *Journal of Sports Sciences*, 25(5), 497–517. <https://doi.org/10.1080/02640410600946779>.
- Wang, T., Ren, X., & Schweizer, K. (2017). Learning and retrieval processes predict fluid intelligence over and above working memory. *Intelligence*, 61, 29–36. <https://doi.org/10.1016/j.intell.2016.12.005>.
- Wiestler, T., & Diedrichsen, J. (2013). Skill learning strengthens cortical representations of motor sequences. e00801 *eLife*, 2. <https://doi.org/10.7554/eLife.00801>.
- Wong, A. L., Lindquist, M. A., Haith, A. M., & Krakauer, J. W. (2015). Explicit knowledge enhances motor vigor and performance: Motivation versus practice in sequence tasks. *Journal of Neurophysiology*, 114(1), 219–232. <https://doi.org/10.1152/jn.00218.2015>.